

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 19:36:23

PAGE 1

REFERENCE NO: 308

This contribution was submitted to the National Science Foundation as part of the NSF CI 2030 planning activity through an NSF Request for Information, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Author Names & Affiliations

- Camille Crittenden - Deputy Director, CITRIS (Center for Information Technology Research in the Interest of Society) and the Banatao Institute, University of California
- Patrick Schmitz - Associate Director, Research IT; Program Director, Berkeley Research Computing

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Crittenden: IT infrastructure and policy for applications of broad social benefit (civic engagement, sustainable infrastructures, health, robotics and artificial intelligence). Schmitz: Research IT, Cyberinfrastructure in support of campus research

Title of Submission

Cyberinfrastructure: Machines, Networks, Software, and People to meet Research needs through 2030

Abstract (maximum ~200 words).

Research universities are pushing the boundaries of current capacity for managing, storing, and analyzing data in fields across the humanities and social sciences, as well as traditional STEM fields. In order to facilitate collaboration and discovery in these data-intensive fields, universities and funding agencies must be prepared to invest in hardware, software, and networking infrastructure, and to train, hire, and promote the staff necessary to support investigators and programs. Demand for data scientists with cross-over domain expertise will continue to grow, and universities are responding by creating new courses, programs, and schools. Institutions of higher education provide the ideal organizational framework not only to develop cutting-edge ideas in cyberinfrastructure but also to cultivate the next generation of experts who will implement them.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

Current research challenges include rapid advances in many aspects of artificial intelligence: machine learning, neural networks, machine vision, deep learning, as well as related applications in virtual and augmented reality.

Computational methods and data-intensive techniques are growing across nearly every domain on campus. For some domains (such as the physical sciences), there are established paths for graduate students to learn and keep pace with the techniques; faculty in these domains are generally prepared to use the associated tools as their research careers begin. However, many other domains are just

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 19:36:23

PAGE 2

REFERENCE NO: 308

adopting the new tools, or are facing a new order of scale of computation and data management, and they lack the background to effectively use the tools. Without on-campus training and support, they have limited capacity to keep up with, much less advance, research in their domains.

Universities are also being challenged to meet demand for data science researchers and are amplifying educational programs in this area. The swelling ranks of students place demands on cyberinfrastructure for their data science classes and research at all levels. As these students progress through introductory courses and into upper division and graduate courses, the distinction between instruction and research blurs; research computing programs must be able to support this broader cohort of users and applications to meet the demand for data science training.

Currently, faculty use NSF funding to purchase computation cycles, storage, and equipment. When improved cyberinfrastructure is available as a campus resource, it is less necessary to create – and budget for – bespoke solutions, and researchers can spend a greater percentage of their research budgets investigating specific research questions. Centralizing campus resources and expertise increases organizational efficiency and reduces demands on local (departmental) IT staff. While regional collaborations can provide some efficiencies provisioning core resources, campus level training and support for researchers are essential to effective use of resources.

The availability of national computation, networking, and storage infrastructure is valuable, but occasionally duplicates or may be seen as in conflict with local campus resources. NSF has an important role to play in supporting campus-level CI resources to ensure early stage research, and preparation of research workflows that can take advantage of national resources. Relatively modest investments at the campus level can provide significant returns on research results. The recent CC*DIBBS conference demonstrated how a range of NSF awards can help innovative cross-disciplinary projects to flourish, that might otherwise not have been implemented.

In addition to local solutions and national resources (e.g., XSEDE), commercial cloud providers may also play a role in the cyber-ecosystem. Cost-benefit analyses vary by campus, and the shifting landscape of economic models among providers and campuses will take some time to settle. Significant questions remain on the role that cloud resources play in the broader research workflow, which can include local sensors or instruments, issues of latency, data mobility vs. computational mobility, and long-term sustainability of data and analytic resources.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Campus cyberinfrastructure must be prepared to support greatly expanding enrollments in computer science classes, especially in entry-level data science courses. Hybrid solutions (e.g., leveraging cloud resources) may apply in some cases, and local or national resources may provide solutions for certain courses; NSF must support ongoing research into scalable, sustainable models to support this kind of instruction.

Federated vs. central model of data storage & IT support: Many research institutions are addressing the long-standing tension between a federated model (in which each lab/center hosts its own computation and data resources and employs its own IT support) vs. a shared/centralized model. At the next organizational level, the University of California overall has also piloted programs starting a decade ago for shared computing infrastructure for clusters of northern and southern UC campuses. We learned from this program that funding capital expenditures and basic systems administration is not sufficient for the long-term success of high-performance computing center, and providing only traditional HPC support fails to meet the research computing needs of the broader community. Greater investment in research IT consulting personnel is essential. Campus leadership are working to better coordinate centers/units (libraries, research centers, etc) and to encourage alliances across disciplinary borders. The Berkeley Institute for Data Science is an example of a hub for connecting researchers in this manner. However, fostering community and collaboration is people-intensive; NSF must support this sort of organizational innovation to provide infrastructure that will better serve the “long tail” of research.

Storage is increasingly problematic as volume, variety, velocity of data increases. We need better tools and protocols for curation (how to decide what to save and what to delete), archiving, and de-accessioning data, as well as for tracking its provenance. Libraries can be partners in developing these plans.

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 19:36:23

PAGE 3

REFERENCE NO: 308

There are associated technical challenges of data locality: should campuses invest in local storage or in fast access to remote storage? An example question from genomics research is whether it is more effective to move data to compute, or vice versa. NSF should support research and development of guidelines and best practices for managing, accessing and analyzing large-scale data sets.

Perhaps the most important need is for NSF to fully recognize that cyberinfrastructure critically includes people. Regardless of how resources for compute and storage are provisioned (campus-HPC, cloud, etc.) researchers depend upon campus-level research consultants/CI engineers/RSEs (Research Software Engineers)/et al. to understand what tools will work best for them, and to help integrate those tools into effective, efficient research workflows. Note that these CI professionals are needed regardless of whether or not the institution has its own infrastructure.

ESnet, part of the Lawrence Berkeley National Lab, has pioneered the concept of science engagement, a model for facilitating research that requires high-speed networking. The NSF-funded Pacific Research Platform builds on this model to enhance capacity in data-intensive domains such as particle physics, astronomy and astrophysics, biomedical sciences, earth sciences, and high-resolution video and visualization. Network engineers, CI engineers and research specialists are prepared to investigate and overcome the physical networking challenges, but assessing research needs and improving workflows also requires engagement specialists, an emerging field that will be critical to accelerating scientific discovery and collaboration.

We continue to look for efficiencies in how we provision and administer core computation, network, and data resources. However, as the community of researchers using computationally intensive and data-intensive methods grows, we must also grow the staff that support them; while some efficiencies are possible through self-help guides, tutorials, etc., researchers need consulting to be effective. Tools and technologies are complex, and constantly evolving, and domain researchers lack the time and training to independently keep abreast of changes in cyber infrastructure technologies. Professional CI staff must be understood as part of the essential CI that supports research. Commonly used F&A models do not reflect the scope of need in this area, and NSF must help to develop models for research grant planning that include support for the essential consulting work that provides access to and integration of CI in support of research.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

Workforce development is an essential consideration. How will people be sourced and trained? The emerging field of cyberinfrastructure engineer is promising – incumbents in these positions are finding each other through professional conferences, meetings, and online platforms. This community-building should be fostered and supported. Consideration should also be given to building the pipeline. What does a successful educational path include?

Cybersecurity is essential for the future of cyberinfrastructure. Tools are needed to de-identify or otherwise protect data. Many researchers do not understand the security implications of using cloud resources. Much of this challenge is actually organizational and policy-related – challenges that must be met at the campus level. However, for campuses that have not addressed this (e.g., those with no medical school that would already have forced institutional support for securing and managing clinical data), it is a significant challenge to bootstrap the support for secure data management, secure analysis enclaves, and related infrastructure. NSF should support programs that provide the start-up funding to develop secure CI programs.

There is a disconnect between research and practice around cyberinfrastructure. Campus practices and policies are not strongly connected to computer science research. Traditionally, NSF is more likely to fund research in computer science than the implementation, application, or establishing the sustainability of approaches proposed in computer science research. Software engineers and network engineers, as examples, are different in training, expertise, and experience from computer scientists; addressing the various CI challenges described herein often requires engineering expertise as much or more than research scientists. We encourage NSF to continue its support to develop engineering capacity, such as the CC* cyberinfrastructure engineers, and to support community collaborations to identify and document sustainable funding models that can support these roles longer term.

Campuses need to address overhead rates associated to services such as cloud compute and storage. As an example, University of Washington has removed the overhead on cloud services, to treat those more like capital expenditure on computing and storage equipment. NSF should consider providing incentives to organizations to make this a more common pattern.

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 19:36:23

PAGE 4

REFERENCE NO: 308

Some in our community have the perception that the NSF budget is heavily weighted towards “big science”, with much less available for the “long tail”, including Social, Behavioral, and Economic Sciences (SBE). NSF and others have recognized that SBE uses as much computation as physical sciences did a decade ago, and this is just one example of how investments for the broader community can have an important impact as more researchers outside of the traditional “big science” domains adopt computational and data-intensive methods. We encourage NSF to look for more opportunities to leverage previous investments in cyberinfrastructure for large projects, to make infrastructure and design patterns more widely available to a broad range of domains and researchers. We strongly encourage continued investment across all domains, including SBE. We also encourage a commitment to supporting a diverse range of CI resources (including, but not limited to, traditional HPC) to address needs across the range of domains supported by NSF.

Consent Statement

- “I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).”
-